



# Approach to classify, separate, and enrich objects in groups using ensemble sorting

Rebecca Turk-MacLeod<sup>a,1</sup>, Alon Henson<sup>a,1</sup>, Marc Rodriguez-Garcia<sup>a</sup>, Graham M. Gibson<sup>b</sup>, Gerardo Aragon Camarasa<sup>a</sup>, Dario Caramelli<sup>a</sup>, Miles J. Padgett<sup>b,2</sup>, and Leroy Cronin<sup>a,2</sup>

<sup>a</sup>School of Chemistry, University of Glasgow, G12 8QQ Glasgow, United Kingdom; and <sup>b</sup>School of Physics and Astronomy, University of Glasgow, G12 8QQ Glasgow, United Kingdom

Edited by David A. Weitz, Harvard University, Cambridge, MA, and approved April 23, 2018 (received for review December 16, 2017)

The sorting of objects into groups is a fundamental operation, critical in the preparation and purification of populations of cells, crystals, beads, or droplets, necessary for research and applications in biology, chemistry, and materials science. Most of the efforts exploring such purification have focused on two areas: the degree of separation and the measurement precision required for effective separation. Conventionally, achieving good separation ultimately requires that the objects are considered one by one (which can be both slow and expensive), and the ability to measure the sorted objects by increasing sensitivity as well as reducing sorting errors. Here we present an approach to sorting that addresses both critical limitations with a scheme that allows us to approach the theoretical limit for the accuracy of sorting decisions. Rather than sorting individual objects, we sort the objects in ensembles, via a set of registers which are then in turn sorted themselves into a second symmetric set of registers in a lossless manner. By repeating this process, we can arrive at high sorting purity with a low set of constraints. We demonstrate both the theory behind this idea and identify the critical parameters (ensemble population and sorting time), and show the utility and robustness of our method with simulations and experimental systems spanning several orders of scale, sorting populations of macroscopic beads and microfluidic droplets. Our method is general in nature and simplifies the sorting process, and thus stands to enhance many different areas of science, such as purification, enrichment of rare objects, and separation of dynamic populations.

ensemble sorting | microfluidics | enrichment | separation | purification

The sorting of objects is a fundamental operation which enables a variety of preparatory techniques such as separation (1, 2), enrichment (3, 4), fractionation (5), and purification (6). Achieving effective sorting is thus of vital importance in many fields including biology (3), chemistry (7), materials science, and engineering (5). Most of the effort in improving sorting techniques centers around two areas: the degree of segmentation (2) and the measurement precision (3). Segmentation refers to the ability to partition the objects into groups (ensembles), down to the ideal limit of being able to manipulate objects on a one-by-one basis. The second area of focus aims to improve the ability to measure the sorted objects by increasing the sensitivity as well as reducing the various errors associated with the desired sorting parameter. Sorting different populations of objects is an exceedingly common operation. It is either used as a preparatory stage before further study or usage (1), or for its own sake (7). Specifically, there is a need to sort such fundamental systems as cells (8) or microfluidic droplets (2, 9) and much effort is focused on improving their separation.

The first target that is often sought is improved segmentation; it would be easiest to classify which desired population to sort objects into by manipulating and evaluating them one by one. For most nonmacro bodies, however, such manipulations are extremely difficult, time-consuming, or in some cases impractical. In fact, it is sometimes impossible to reach one-by-one sorting due to the

evaluation criteria requiring a larger number of objects (i.e., measuring affinity or interaction between objects). The second common target is improving the precision of the sorting decision. Reducing the error rate in the sorting decision is an obvious goal. The higher the certainty in the sorting decision, the higher the achievable theoretical separation of the different target populations.

We present here a scheme that allows for sorting over different types of objects and over various orders of size while avoiding the need for segmentation of the objects to individuals. Extremely high degrees of separation can be achieved in our scheme up to the threshold of the measurement precision. Furthermore, our method not only withstands high variability in sorting parameters, it can even be enhanced by them. Our robust method relies upon ensemble sorting. We intentionally allow the sorting decision to be made on an ensemble of objects, with the trade-off being the need for cycles of sorting with the purification increasing as the number of cycles increases. The distance from the ideal segmentation of one object at a time can be offset by increasing the number of sorting cycles. The proposed scheme is conceptually simple and can be implemented in a variety of scales and types of object and for a variety of desired outcomes. We provide two proof-of-concept systems that operate on different scales.

## Results

**Sorting Scheme.** The scheme is depicted in Fig. 1 for the case of two object features, blue and red. An initial population containing

## Significance

The sorting and separation of objects, from single cells, droplets, or macromolecules, to materials, is an important operation central to many scientific investigations. Indeed, it is central to the fabrication of machines and detection technologies; however, most methods focus on separating individual components which can be costly and time-consuming, and it is difficult to achieve high throughput at low error rates. Here, we propose a simple method applicable over a large size and material range to sort objects in groups or ensembles using a set of registers, which are themselves sorted again, and then re-sorted iteratively. The result is a lossless enrichment process that leads to a very high purity and can cope with errors and highly dynamic populations.

Author contributions: M.J.P. and L.C. designed research; R.T.-M., A.H., and M.R.-G. performed research; G.M.G., G.A.C., and D.C. contributed new reagents/analytic tools; R.T.-M., A.H., and M.R.-G. analyzed data; D.C. made the concept video and picture renderings of the device; and R.T.-M., A.H., and L.C. wrote the paper.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>R.T.-M. and A.H. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: Miles.Padgett@glasgow.ac.uk or Lee.Cronin@glasgow.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1721929115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1721929115/-DCSupplemental).

Published online May 14, 2018.

a mixture of both types of object is introduced to the system and fed into one of four registers, or bins—for example, the upper left-hand “A” register. A small group, or ensemble, of the objects is then sent into the central module for viewing and decision-making. Here, if the majority in the ensemble is blue it is sent to blue register “B,” and if red, it is sent to red register B. Once the initially loaded A register is empty, one of the B registers is sorted one ensemble at a time into the blue and red A registers. Following that, the second B register is similarly sorted. The registers are continuously sorted one by one until the stopping criterion (i.e., a preestablished degree of separation) is met and the separated populations are then sent from their respective bins to the output. Here we define degree of separation, or purity, as the fraction of correctly sorted objects/total objects in the corresponding registers (e.g., the number of blue objects/the total number of objects in the blue registers of A and B).

**The Theory of Ensemble Sorting: Principles of Operation.** To understand the properties of ensemble sorting, and for experiment planning, we first define the relationship between the different sorting parameters. Eq. 1 shows the duration of time to sort a complete population using the ensemble sorting scheme.

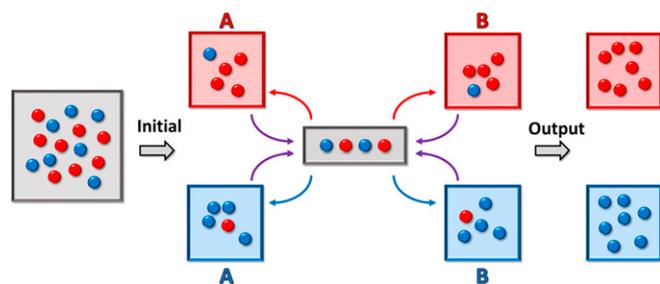
$$T_{\text{ensemble}} = \frac{N}{S} \cdot E \cdot (t_h + t_d), \quad [1]$$

where  $T_{\text{ensemble}}$  is the total sorting duration,  $N$  is the number of cycles needed to accomplish the desired separation outcome,  $S$  is the average ensemble size,  $E$  is the total number of objects,  $t_h$  (equivalent to  $t_{h1} + t_{h2}$ ) is the duration of physical handling where several objects from a population are gathered into an ensemble ( $t_{h1}$ ) and moved into the targeted population ( $t_{h2}$ ), and  $t_d$  is the decision time for sorting. For most physical systems  $t_h \gg t_d$ .  $N$  can either be predetermined by the user, or estimated by simulation, given the starting ratio and desired separation value (Fig. 2).

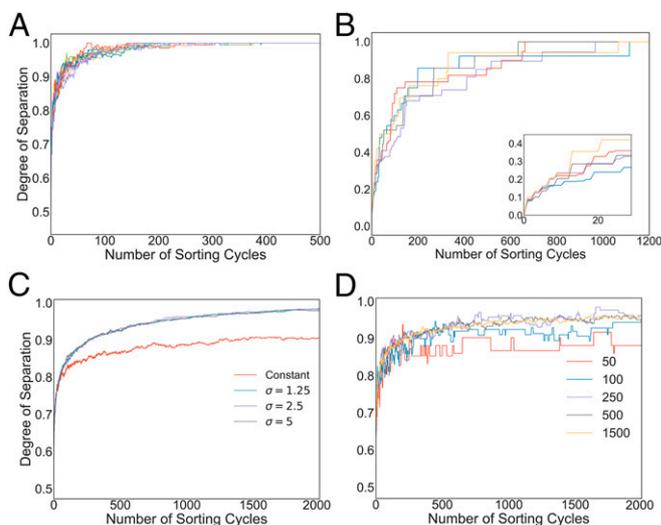
A special case arises when the state of the objects in the population is dynamic. Over a time scale defined as  $t_{\text{dyn}}$ , the objects can change their state, for example from diffusion of compounds, chemical reaction/decomposition, or stochastic processes. If we define the amount of time for one cycle ( $N = 1$ ) of ensemble sorting as

$$T_{\text{cycle}} = \frac{E}{S} \cdot (t_h + t_d). \quad [2]$$

This leads to three scenarios for the dynamic system; the first is the case where  $T_{\text{cycle}} \ll t_{\text{dyn}}$ . In this case we can essentially ignore the dynamics of the states in the population as the separation process will perform efficient sorting irrespective of the dynamics of the objects. The opposite case, where  $T_{\text{cycle}} \gg t_{\text{dyn}}$ , leads to a



**Fig. 1.** Scheme of the ensemble sorter. The sorter consists of four registers, one blue and one red for both A and B. Also shown is the measurement area in the middle, which contains an ensemble to be evaluated before the sorting decision is made. Purple arrows indicate input to the measurement area, and red and blue arrows indicate output to the corresponding registers. See *SI Appendix, Fig. S1* and *Movie S1* for step-by-step visualizations.



**Fig. 2.** Simulation plots for sorting, enrichment, and several ensemble-size distributions. (A) Separation of populations starting with a ratio of 1:1. (B) Enrichment from a starting ratio of  $\sim 1:1,000$ ; (Inset) Enlargement of the initial portion of the enrichment simulation. (C) Different ensemble-size distributions, both constant and normal with varying SD. (D) Separation using several different ensemble sizes.

situation where the system will only be able to separate the population if  $t_{\text{dyn}}$  is not constant, but rather is positively correlated with separation; i.e., if each object is influenced by its neighbors in a concentration-dependent manner. Even if this is not the case, the user may be able to adjust the parameters  $E$  and  $S$  to generate a more favorable scenario. Lastly, the case where both time scales are of the same order of magnitude still allows for separation, yet at a cost. The state dynamic will decrease the efficiency of separation, yet repeated sorting cycles will overcome this dynamic hindrance, at the expense of time required to achieve separation. To evaluate sorting under different conditions, we can define sorting efficiency as the odds of improving the separation from the starting situation. There are two opposing effects to consider. The first is that if the population is close to 50% separation, most ensembles will be relatively evenly mixed and effective separation will be slow, at least initially. The second effect is that when the population ratio is heavily skewed, there is a larger chance of generating ensembles that are of only one type. These ensembles do not change the level of overall separation in the system, but they do cost time and added effort. The conclusion is that although sorting is most efficient for a population starting far from 50%, it comes at a cost of many ensembles being needlessly sorted.

Sorting by ensembles rather than by individual objects leads to sorting results that after one cycle are at a relatively low separation value. However, the key attribute of the system is that even though each sorting cycle would on average be worse than an equivalent sorting on an individual basis, the repetition of sorting cycles will improve the resulting separation. Since the distribution of objects in any single ensemble is random, there may be instances where after a cycle of sorting, the overall separation decreases. However, on average the separation will follow a clear rising trend. To examine this ensemble-sorting scheme, we developed a simulation that was used to evaluate the performance under different operating conditions. We can see in Fig. 2A an example of several experiments run under identical conditions. The starting ratio between two classes of objects was set to 1:1 and the simulation was run for a predetermined number of cycles. As expected, the stochastic behavior embedded into the simulation leads to variation in the separation ratio along the

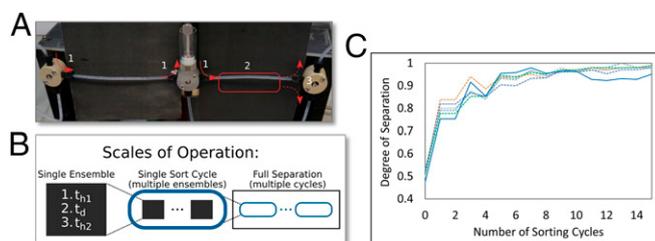
course of the sorting cycles. The rising trend is monotonic and follows an exponential behavior.

When we instead start from a very low fraction of one type, well under 1%, successive cycles of simulated sorting quickly increase the proportion of the desired object class by orders of magnitude (Fig. 2B). Using these simulations, we have validated the ensemble-sorting scheme for use under various operation requirements and over many different orders of scale for the overall number of objects, the size of the ensemble, and the initial population ratio.

An important parameter in the sorting behavior is the size of the ensemble. In Fig. 2C we show simulations using different distributions of ensemble size. Along with a constant ensemble size, we also ran normal distributions with varying SD. The lowest rate of improvement is obtained with constant ensemble size, whereas all normal distributions whose mean value is the same as the constant show better performance. Among these normal distributions, the rate of separation improves with the width of the distribution; in other words, the higher the variability of the distribution, the higher the rate of improvement. The reason for this is that the rate is mostly influenced by the smallest ensemble size that is used. Although the normal distributions are symmetric, meaning that on average the ensemble sizes below the mean and those above the mean are the same, occasionally having smaller ensembles has a higher positive impact. The amount of gain from a small ensemble compared with the mean is higher than the loss of a larger ensemble compared with the mean. The conclusion from this is that a simple way to improve the sorting performance is to have a distribution of ensemble sizes that has a high degree of variation. When we vary the ensemble size, we see that this does not change the rate of separation, yet it does affect the variability (Fig. 2D). As the size of the ensemble decreases, the fluctuation in the separation increases.

**Ensemble Sorting in a Macroscopic System.** To test the ensemble-sorting scheme on a physical system, we sorted two populations of millimeter-scale polyethylene beads in water in a liquid handling system with four chambers interconnected with plastic transparent tubing. Both beads had the same physical characteristics such as density, size, and material. The distinguishing factor between them was only their color; one population was pink and the other yellow. To manipulate the beads we used TriContinent syringe pumps. The control principle of the system was stop flow where an ensemble of beads was taken from one chamber and pushed to an observation region. A camera would take a picture of the ensemble and an in-house-developed algorithm would perform an image recognition operation followed by a decision of where to send the beads. For example, if there were more pink than yellow beads in an ensemble, the ensemble would be sent to the pink chamber. If there were an equal number of pink and yellow beads, the system would randomly decide which chamber to send the ensemble into. With the decision made, the pump would push the ensemble of beads into the chosen chamber (Fig. 3). Once the chamber was empty, the process would repeat, in the opposite direction; beads that had just been sorted into the “pink” chamber would be sorted again, followed by beads from the “yellow” chamber (SI Appendix, Fig. S1).

The bead-sorting system was able to reliably perform the ensemble-sorting scheme. A typical experiment involved sorting of several hundreds of beads with an initial distribution equal between the two classes of beads. All of the beads would start in one chamber and then the system would sort each chamber, one at a time, increasing the degree of separation between the two bead populations. The result of a separation experiment can be seen in Fig. 3, where the ratio of both bead populations quickly rises with sorting cycles. In addition to the physical experiments, we also performed simulations of the bead-sorting system using



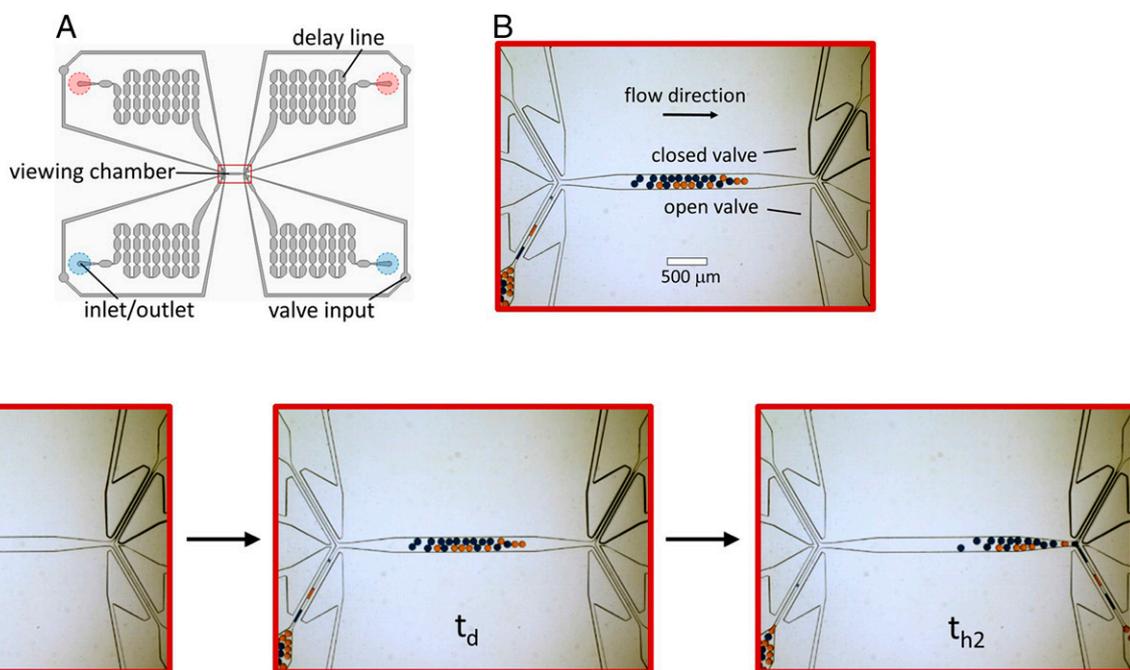
**Fig. 3.** (A) Bead sorter picture and (B) scheme showing the three action scales needed for a full separation. In this cycle, beads move through line (1), are imaged in line (2), and are sorted with valve (3). Refer to SI Appendix for further details about experimental setup. (C) Results from macroscopic bead sorting. Solid line indicates experimental data, dashed lines indicate simulated data based on ensemble frequencies from experiment.

similar experimental parameters. As can be seen in Fig. 3, the simulations show a similar behavior to that seen in the physical system.

**Ensemble Sorting in a Microfluidic System.** We then developed a microfluidic system to test ensemble sorting of micrometer-scale objects, in this case aqueous droplets dispersed in fluorinated oil. Such a system could be adapted to work on many different sorting criteria, thereby having a broad usage case. Two populations of red and blue droplets of  $\sim 80\text{-}\mu\text{m}$  diameter were generated and loaded into a microfluidic ensemble sorter device (Fig. 4). This device is composed of four symmetric modules of inlets/outlets, delay lines, and high-aspect-ratio valves (10), a viewing chamber, and an additional isotonic aqueous reservoir layer (11) to mitigate droplet evaporation (12). Inlets and outlets led to three-dimensional tapered cavities within the device that served as bins for droplet trapping and incubation between modules (SI Appendix). The device and sorting mechanism were analogous to that of the macroscopic system; droplets started in one bin within the microfluidic chip, were pushed through a delay line by syringe pumps, and were sorted into respective “red” or “blue” bins, with their direction controlled by pressure-actuated valves (Movies S1 and S2). After a set number of “empty” pushes were performed, the starting bin was determined to be empty, and the next cycle (sorting from the next bin) was initiated. Control of the pumps and valves, as well as image analysis, was automated with a custom made LabVIEW virtual instrument (National Instruments).

The microfluidic system was capable of sorting red and blue droplets at an initial ratio of 1:1 to an extent comparable to that of a corresponding simulation (Fig. 5A). Separation is initially rapid, slowly trailing off and reaching over 90% at  $\sim 10$  cycles. Sources of experimental error include timing of droplet recognition ( $t_d$ ), sluggish exit of droplets from the inlet reservoirs ( $t_{h1}$ ), and occasional variation in resistance between outlets ( $t_{h2}$ ). An important component of the device is the aqueous reservoir layer, without which droplets shrink over time and become unrecognizable by the image recognition algorithm (SI Appendix, Fig. S14). Notably, by design of the microfluidic sorter, a distribution of ensemble sizes is a natural phenomenon of the system, leading to enhanced sorting as predicted by simulation. During a sorting module, the droplets are initially highly packed, and become less packed as they move through the delay line (SI Appendix, Fig. S24). The geometric design of the delay line also facilitates mixing of the droplets before and after sorting, ensuring population shuffling, and the transition time between modules offers an additional opportunity for controlled cocultivation of the semisorted populations in their respective bins.

We then endeavored to enrich a population of red microdroplets (1%) in a background of blue microdroplets. To accomplish this, we implemented a “simple” decision-making



**Fig. 4.** Microfluidic ensemble sorter. (A) Two-dimensional schematic of sorter layer. Red and blue circles indicate integrated bins. (B) Microscopic image of sorter during operation. (C) Images of microfluidic sorter operation during handling times  $t_h$  and decision time  $t_d$ .

scheme in the sorting algorithm, which would initially sort any ensemble containing one or more red droplets into the red module (*SI Appendix*, Fig. S21 and *Movie S3*). This scheme is similar to previous work enriching rare cells, but further enriches with repeated sorting cycles, and does not require a filtration step (13). Using the simple sorting scheme, the red register was populated with 26% red droplets after five cycles. At that point, the program switched to the “complex” decision-making scheme that was implemented in the previous (1:1 ratio) experiments, after which the red register was populated with ~100% red droplets after a further 11 cycles (Fig. 5B).

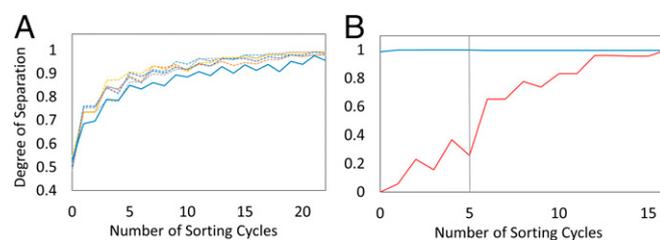
### Discussion

We have used simulations as well as physical systems to implement the ensemble-sorting scheme. The results show that the scheme works for different modes of operation such as separation, where the goal is to effectively separate an initial mixed population of objects, and enrichment, where a mixed population has one desirable class that initially is at a low ratio. Other usages include fractionation and purification, which on a technical level are the same as separation and enrichment, respectively. By its nature our system allows the user to adapt and refine sorting criteria in real time, even when starting with a population of dynamic and heterogeneous objects. The ensemble-sorting scheme is thus able to undertake different modes of operation for different purposes with only mild modifications to the algorithm.

Our microfluidic design allows for the separation of a range of population sizes, and could be easily modified (by changing the length of the delay line, optimizing input velocities) to accommodate experimental variations. One such example would be sorting cells by a dynamic characteristic such as their ability to intake certain molecules from solution. In addition, one can change the algorithm decision threshold (the point at which to change from simple to complex) in real time to efficiently enrich a minority population at a range of starting values. The concept can be applied to any system where there is a measurement capable of distinguishing between different classes, even if the distinction is poor. It would also work on any size scale from

molecules to cells, droplets, and even macrosized objects. Furthermore, our system is amenable to other real-time variations, such as changes in environmental parameters, which could be easily programmed and integrated into the sorting scheme. This could be useful in *in vitro* selection/evolution experiments where progressively more stringent selection pressure is applied throughout sorting cycles. In addition, different solvents/populations of interacting objects can be introduced to the system at any time, and withdrawal of objects of interest/removal of waste objects can also be performed without disruption to the rest of the system.

Our device fabrication materials (polydimethylsiloxane, PDMS) and hardware setup (light microscope and syringe pumps) are relatively low-cost (as they are flexible and can be put to various uses), easily available, and interchangeable. Our image recognition/sorting criteria are not limited to colorimetric detection, but rather could incorporate other visual or physical effects on micrometer-scale objects. For example, it has been shown that microfluidic droplets can change size and physical properties depending on their chemical contents/reaction progress (14), and that some compounds can exchange between droplets (15). Therefore, our system could be used to sort populations



**Fig. 5.** Sorting efficiency of microfluidic ensemble sorter. (A) Sorting from a starting ratio of 1:1. Solid line indicates experimental data for one population (blue droplets), dashed lines indicate simulated data based on ensemble frequencies from experiment. (B) Enrichment of 1% red droplets (red line) in a background of blue droplets (blue line). Vertical line indicates point where decision algorithm changed from simple to complex.

